

André Araújo Luchine

**Estimando a relação entre as habilidades
cognitivas da Educação Básica e o salário
futuro**

Brasília

2015

André Araújo Luchine

Estimando a relação entre as habilidades cognitivas da Educação Básica e o salário futuro

Monografia apresentada junto ao curso de Ciências Econômicas da Universidade de Brasília como requisito parcial à obtenção do título de Bacharel.

Universidade de Brasília – UnB
Faculdade de Economia, Administração e Contabilidade
Departamento de Economia

Orientadora: Ana Carolina Pereira Zoghbi

Brasília
2015

André Araújo Luchine

Estimando a relação entre as habilidades cognitivas da Educação Básica e o salário futuro

Monografia apresentada junto ao curso de Ciências Econômicas da Universidade de Brasília como requisito parcial à obtenção do título de Bacharel.

Trabalho aprovado. Brasília, 15 de julho de 2015:

Prof^a. Dr^a. Ana Carolina Pereira Zoghbi
Orientadora

Prof. Dr. Rafael Terra de Menezes
Banca Examinadora

Brasília
2015

À minha família e aos meus amigos.

Resumo

O objetivo do trabalho é mostrar a relação entre a qualidade da educação básica, expresso pela nota da prova objetiva do Exame Nacional de Ensino Médio (ENEM) de 2006, e o salário futuro, expresso pelo rendimento médio de 2011 segundo dados da Relação Anual de Informações Sociais (RAIS). Segue-se a metodologia de Curi e Menezes-Filho (2005), com adaptações, devido à diferença de estrutura do banco de dados, que em vez de agrupar indivíduos em células de acordo com as características ao longo do tempo, observa-se dados defasados de diferentes períodos para mesmo indivíduo, e a utilização de variáveis instrumentais, para melhor captar o efeito de uma variação exógena de nota sobre salário futuro. Conclui-se que não há indícios suficientes de um efeito significativo da qualidade da educação sobre os rendimentos, que contraria a literatura, mas que pode ser explicado pelo intervalo limitado entre a conclusão do ciclo da Educação Básica e a inserção no mercado de trabalho, viesando os resultados.

Palavras-chaves: Educação Básica. Qualidade da educação. Salário. Variáveis Instrumentais.

Abstract

The aim of the work is to show the relationship between educational quality, expressed by the score of the ENEM 2006 standardized test, and the future wage, expressed by the average earning according to the RAIS 2011 database. The methodology utilised from the paper of Curi and Menezes-Filho (2005) is adapted due to data structure differences that match the individual among different databases in lagged periods instead of clustering observations based on similar characteristics across time. Also, this work utilises an instrumental variable approach to capture the effect of an exogenous change of the score on wages. There is not enough evidence of a significant effect of educational quality on earnings, countering the existing literature. However, this can be explained due to a small gap between finishing high school and entering the labor force, which produces biased results.

Keywords: Basic education. Educational quality. Wage. Instrumental Variables.

Lista de ilustrações

Figura 1 – Gráfico comparativo da média do logaritmo do salário médio real por UF	18
Figura 2 – Gráfico comparativo da média da nota da prova objetiva do ENEM por UF	19
Figura 3 – Gráfico comparativo do percentual de migrantes por UF	20

Lista de tabelas

Tabela 1 – Estatística descritiva do salário médio real de 2011	17
Tabela 2 – Estatística descritiva da nota da prova objetiva do ENEM 2006 . . .	19
Tabela 3 – Características da amostra sobre composição geográfica e migração	20
Tabela 4 – Estimação dos modelos originais	30
Tabela 5 – Estimação dos modelos derivados #1	31
Tabela 6 – Estimação dos modelos derivados #2	32
Tabela 7 – Estimação dos modelos derivados #3	33
Tabela 8 – Estimação dos modelos derivados #4	34
Tabela 9 – Estatísticas descritivas das variáveis, por geração	36

Lista de abreviaturas e siglas

DATASUS	Departamento de Informática do Sistema Único de Saúde
ENEM	Exame Nacional de Ensino Médio
IV	Variável Instrumental
MQ2E	Mínimos Quadrados em 2 Estágios
MQO	Mínimos Quadrados Ordinários
RAIS	Relação Anual de Informações Sociais
UF	Unidade da Federação
UNESCO	Organização das Nações Unidas para Educação, Ciência e Cultura

Sumário

Introdução	10
I Referencial teórico	11
1 Revisão de literatura	12
II O modelo e seus resultados	16
2 Dados e metodologia	17
2.1 Análise dos dados	17
2.2 Metodologia aplicada	22
2.2.1 Variáveis Instrumentais	22
3 Discussão e análise dos resultados	24
Conclusão	25
Referências	26
Apêndices	28
APÊNDICE A Regressões	29
APÊNDICE B Estatísticas descritivas das variáveis	35

Introdução

Nas últimas décadas, a literatura de Economia da Educação ganhou um enfoque na qualidade da educação e sua relação com o salário. Muitos estudos feitos no Brasil abordam o tema, porém o arranjo dos dados não permite trabalhar com metodologias mais robustas. Este trabalho visa a apresentar o efeito da qualidade da Educação Básica sobre os salários futuros utilizando uma metodologia de dados e abordagens econométricas diferente.

O ponto de partida do trabalho é o estudo de Curi e Menezes-Filho (2005), que utiliza dados da Pesquisa Nacional de Amostra por Domicílio (PNAD) de 1982, 1995, 1997, 2001 e 2003 e do Sistema de Avaliação da Educação Básica (SAEB) de 1995 e 1997 para agregar as observações em grupos de 10 ou mais indivíduos de acordo com sexo, raça e UF para um mesmo ano e depois empilhar os diversos cortes transversais em um pseudopainel, estimando o modelo por MQO.

Com isso, a metodologia aplicada ao longo deste trabalho sofre algumas alterações em relação à metodologia aplicada em Curi e Menezes-Filho (2005). Primeiramente, ela permite identificar o indivíduo em diferentes bases de dados e juntar a informação das variáveis em períodos de tempo distintos, como um *cross-section* de dados defasados. Isso é importante para que se consiga observar o desempenho no ENEM 2006 e o salário médio de 2011 do mesmo indivíduo, informações cruciais para o modelo.

Outra mudança importante se dá na abordagem econométrica, estimando o modelo por MQO e variáveis instrumentais (MQ2E), de modo a observar o efeito de uma variação exógena da qualidade da educação sobre os salários. Toda a metodologia adotada é explicitada no capítulo de Dados e Metodologia. Por fim, no capítulo de Discussões reportam-se os resultados obtidos dos modelos MQO e MQ2E.

Parte I

Referencial teórico

1 Revisão de literatura

A literatura acerca da Economia da Educação possui raízes na Teoria do Crescimento e Desenvolvimento Econômico, uma vez que há curiosidade acerca dos fatores os quais levavam a economia de um país a crescer e se desenvolver. Para isso, é preciso falar primeiro dos modelos de crescimento econômico e como essa relação com a educação se deu.

Posterior ao modelo de crescimento desenvolvido por Roy Harrod e Evsay Dornar em 1946, o modelo de Solow–Swan, mais conhecido como modelo de Solow, ganhou notoriedade ao ser publicado em 1956, sendo um dos mais influentes modelos de crescimento econômico exógeno. Solow (1956) trabalha com uma função de produção aparentemente simples, de formato Cobb-Douglas, dependendo apenas de capital físico, determinado por K , e força de trabalho, determinado por L . O fator de produtividade, também denominado “eficiência do trabalho” ou “progresso tecnológico”, é determinado de forma exógena e afeta apenas a força de trabalho.

Para Solow, as taxas constantes de crescimento da força de trabalho e do progresso tecnológico, bem como a depreciação constante do capital físico implicam um crescimento no longo prazo alicerçado pelo progresso técnico, uma vez que ele mitiga os efeitos de retornos decrescentes de K .

Porém havia mais do que somente capital físico e força de trabalho para explicar o crescimento de uma economia. Jones (2000) atesta que Solow, em seu trabalho “*Technical Change and the Aggregate Production Function*”, datado de 1957, explora a existência de um fator cuja taxa de crescimento afeta a taxa de crescimento da economia de forma perfeitamente linear e que transborda o efeito ponderado da taxa de crescimento de K e L . Solow denominou esse fator como “medida da nossa ignorância”, que também ficou conhecido por “resíduo de Solow” e mais comum na literatura de Crescimento Econômico, a “produtividade total dos fatores”.

A atenção se volta para um possível elemento explanador do resíduo de Solow: capital humano. Ao final da década de 50 e início da década de 60, trabalhos de Jacob Mincer, em 1958, Theodore Schultz, em 1960 e Gary Becker, em especial, em 1964, abordariam em primeira mão sobre a questão de explorar o conceito de capital humano e entender sua dinâmica com o crescimento econômico.

Mincer (1958) define capital humano como qualificação educacional ou profissional adquirida e trabalhada que agrega produtividade à força de trabalho. Mincer também procura compreender a razão do investimento em capital humano e da escolha de investir agregadas por diferentes níveis: escolaridade, idade e tipo de profissão

exercida. A conclusão é de que uma correlação positiva entre capital humano e produtividade é encontrada, visto que a produtividade é mensurada pelos rendimentos, nesse caso.

A intuição sobre a definição do capital humano é corroborado por Schultz (1960), que argumenta que o capital humano é um investimento no indivíduo que gera aumento de produtividade, resultando, no agregado, em crescimento do produto do país. Capital humano, portanto, é objeto de delicada análise para a magnitude de crescimento de uma economia.

A análise de Becker, em seu livro “*Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*”, de 1964, traça uma das primeiras relações entre capital humano, salários e educação. A escolaridade do indivíduo sinaliza um certo nível de capital humano, agindo como uma *proxy*, e tem um efeito positivo sobre a remuneração. Mincer (1984) encontra evidências que corroboram com a visão de Becker.

Até a década de 90, a literatura de Economia da Educação possuiu um enfoque em quantidade de educação como *proxy* de capital humano, mas desde então outras *proxies* são sugeridas. Barro (1991) sugere *proxies* voltadas para refletir a qualidade da educação, e não a quantidade. Nesse estudo, a tentativa de mensuração da qualidade da educação se deu por meio da fração discente–docente. Barro aponta uma correlação negativa entre essa *proxy* da qualidade da educação com a taxa de crescimento da renda *per capita*.

Hanushek (2013) argumenta que a qualidade da educação, ao representar o estoque de capital humano, pode ser traduzido em produtividade e crescimento econômico, especialmente em países com taxas de crescimento baixas, pois o efeito é relativamente maior. Tal conclusão corrobora o modelo teórico de Romer (1990).

A preocupação em atingir um nível de estoque de capital humano já era levada em consideração 10 anos antes de Hanushek tratar dessa argumentação, pois seu estudo considerou que políticas educacionais deveriam ser mais qualitativas, a fim de aumentar a qualidade da educação, e não quantitativas (HANUSHEK, 2003).

Murnane, Willet e Levy (1995) chegam a sugerir que a abordagem seja restrita apenas à qualidade da educação, uma vez que em seu modelo ela possui efeito marginal superior ao de anos de estudo, fazendo com que a variável sequer seja significativa.

Essa insignificância da quantidade da educação também é corroborada por Barro e Lee (2001), em que comparou-se dados de 39 países durante o período de 1960 a 2000 para testar a correlação entre o resultado dos testes padronizados internacionais, uma *proxy* para qualidade da educação, e quantidade de educação, medida

em anos de estudo. Barro encontrou uma correlação positiva, porém um efeito marginal estatisticamente insignificante de anos de estudo sobre desempenho nos testes.

O desempenho escolar em testes padronizados se tornou *proxy* predominante de qualidade da educação na literatura. Os trabalhos de Psacharopoulos e Woodhall (1985), Hanushek (1986), Bishop (1989) e Hanushek e Kimko (2000) encontram boas evidências que sustentam uma correlação boa com a qualidade da educação a ponto de poder ser quantificada.

Hanushek e Kimko (2000) fazem uma análise comparativa entre as taxas de crescimento da economia e a qualidade da força de trabalho, mensurada a partir dos resultados de desempenho em testes padronizados de matemática e ciências realizados pela UNESCO.

Sua análise detecta uma grande sensibilidade dos países asiáticos à qualidade da educação, fazendo com que os retornos no crescimento econômico sejam maiores que dos demais. A diferença medida por variáveis *dummy* de continente revelam que entre países de diferentes continentes a diferença do efeito marginal pode ser de até 10%, enquanto as diferenças entre países do mesmo continente são estatisticamente insignificantes para países asiáticos e africanos.

Os autores concluem que essas diferenças de qualidade da força de trabalho observadas derivam da qualidade das escolas, que influenciam diretamente no desempenho, e, conseqüentemente, impactam as taxas de produtividade e de crescimento.

A endogeneidade da qualidade da educação levou os economistas a procurarem não somente suas *proxies*, mas também seus determinantes. O *background* familiar (variáveis que captam a condição socioeconômica da família e o nível de escolaridade dos pais) aparece nos trabalhos de Murnane, Willet e Levy (1995) e Barro e Lee (2001) como um dos primeiros determinantes da qualidade da educação.

Os indicadores escolares, conhecidos também como *school inputs*, foram objeto de estudo de Felício e Fernandes (2005) para determinar seu impacto na qualidade da educação (“efeito escola”). Os autores tentam capturar o efeito causal dos indicadores escolares sobre a qualidade da educação utilizando dados da prova do Sistema de Avaliação da Educação Básica (SAEB) para alunos do Ensino Fundamental I e II do Estado de São Paulo.

O resultado encontrado, depois de controlado para o efeito do *background* familiar sobre a qualidade da educação, é de que a diferença entre os indicadores escolares de duas instituições de ensino diferentes, explique no máximo 28,4% das notas em Língua Portuguesa e de 8,7% a 34,4% das notas em Matemática.

O trabalho conclue que os indicadores escolares possuem uma parcela de ex-

plicação pequena sobre a qualidade da educação mensurada pelo desempenho escolar, uma vez que foram controladas pelo *background* familiar, que possui um efeito maior e significativo. Por fim, corrobora-se o argumento de Hanushek (2003) de que as políticas educacionais devem estar voltadas para o desempenho escolar.

Estudos relativamente novos na literatura sobre a relação da pré-escola, qualidade da educação e rendimentos salariais sugerem que cursar o ensino infantil influencia a qualidade da educação e a probabilidade de concluir os ciclos escolares (CURI; MENEZES-FILHO, 2009).

Para os autores, o fato do indivíduo ter frequentado a pré-escola também está positivamente correlacionado com o desempenho escolar, representado por um aumento em torno de 7,5% para estudantes da quarta série do Ensino Fundamental e 1% para estudantes do Ensino Médio. Já Chetty *et al.* (2011) encontram que o impacto do ensino infantil é positivo sobre salários, com um aumento em torno de 6,9% até os 27 anos de idade do indivíduo.

Com um enfoque próximo ao *school inputs*, os modelos de *teacher value-added* mostram uma nova *proxy* da qualidade da educação ao trabalhar com variáveis do ambiente escolar. Os trabalhos de Chetty, Friedman e Rockoff (2014a) e Chetty, Friedman e Rockoff (2014b) demonstram que o estimador de *teacher value-added* é consistente para se realizar estimativas da qualidade da educação e proficiência dos alunos.

Depois, os autores trabalham para verificar empiricamente o retorno salarial de estudantes de Washington, DC e Los Angeles, trabalhando com um experimento aleatório a fim de controlar os efeitos não-observados e extrair o efeito causal do *teacher value-added* sobre a nota dos testes de proficiência padronizados dos alunos. A conclusão é de que um professor com um *value-added* 5% maior pode levar um aluno a ter um salário anual de 250 mil dólares.

Parte II

O modelo e seus resultados

2 Dados e metodologia

2.1 Análise dos dados

A base de dados utilizada é um *cross-section* de indivíduos de duas gerações, em que a primeira geração é representada por indivíduos nascidos entre 1984 e 1986 e a segunda geração, entre 1987 e 1989. Em 2006, na época do ENEM, os indivíduos da primeira geração têm idade entre 20 e 22 anos e os da segunda geração, entre 17 e 19 anos. Em 2011, quando são observados no mercado de trabalho, a primeira geração conta com indivíduos de idade entre 25 e 27 anos e a segunda geração, entre 22 e 24 anos. Essa abordagem geracional permite observar os efeitos do atraso educacional no Brasil.

Os dados são obtidos através do ENEM 2006, do Censo Escolar 2006, das RAIS 1984–1994, da RAIS 2006, da RAIS 2011 e do DATASUS.

O banco de dados do ENEM 2006 fornece informações sobre o indivíduo (sexo, cor, município de residência e município onde estuda) e sua nota da prova objetiva e da redação, bem como sua situação socioeconômica, seu *background* familiar e sua percepção sobre o ensino na escola em que estuda (ou estudou) e sobre seus professores. O Censo Escolar 2006 traz informações sobre a infraestrutura das escolas e a RAIS 2006 traz informações sobre escolaridade e salário dos professores e diretores de Ensino Médio do município em que o aluno estuda.

Os dados da RAIS 2011 fornecem os salários médios reais dos indivíduos, a variável central dos modelos. Por fim, os dados das RAIS 1984–1994 e do DATASUS fornecem informações sobre número e salário médio real de professores de Ensino Infantil e sobre a população infantil entre 0 e 5 anos de idade.

Na tabela 1 abaixo observa-se a estatística descritiva do salário médio real de 2011 para as duas gerações. Enquanto há uma grande diferença entre os valores máximos, a média, o desvio-padrão e a mediana permanecem praticamente iguais.

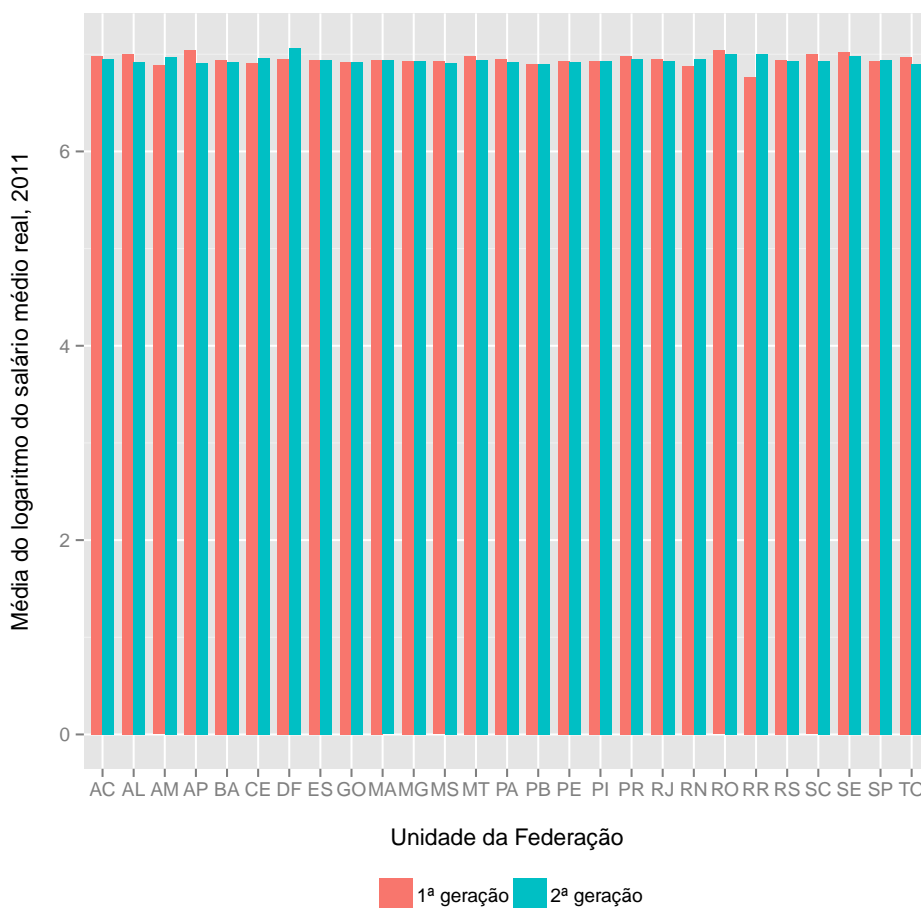
Tabela 1 – Estatística descritiva do salário médio real de 2011

	Salário médio real (2011, R\$)				
	Obs	Mínimo	Máximo	Média / Desvio-padrão	Mediana
1ª geração	11885	0	26.230	1.211,52 / 998,71	933,18
2ª geração	53027	0	40.500	1.215,11 / 1038,41	921,74

Fonte: elaboração própria

Na figura 1, a média do logaritmo do salários é comparada por UF e geração. Observa-se pouca diferença geográfica e geracional, em que os casos mais acentuados ocorrem no Amapá, em Roraima e no Distrito Federal. O maior valor registrado para a primeira geração é no Amapá (7,04) e para a segunda geração é no Distrito Federal (7,07).

Figura 1 – Gráfico comparativo da média do logaritmo do salário médio real por UF



Fonte: elaboração própria

A estatística descritiva da nota da prova objetiva do ENEM 2006 é apresentada na tabela 2 abaixo. Praticamente não há diferença entre as gerações, assim como no caso dos salários.

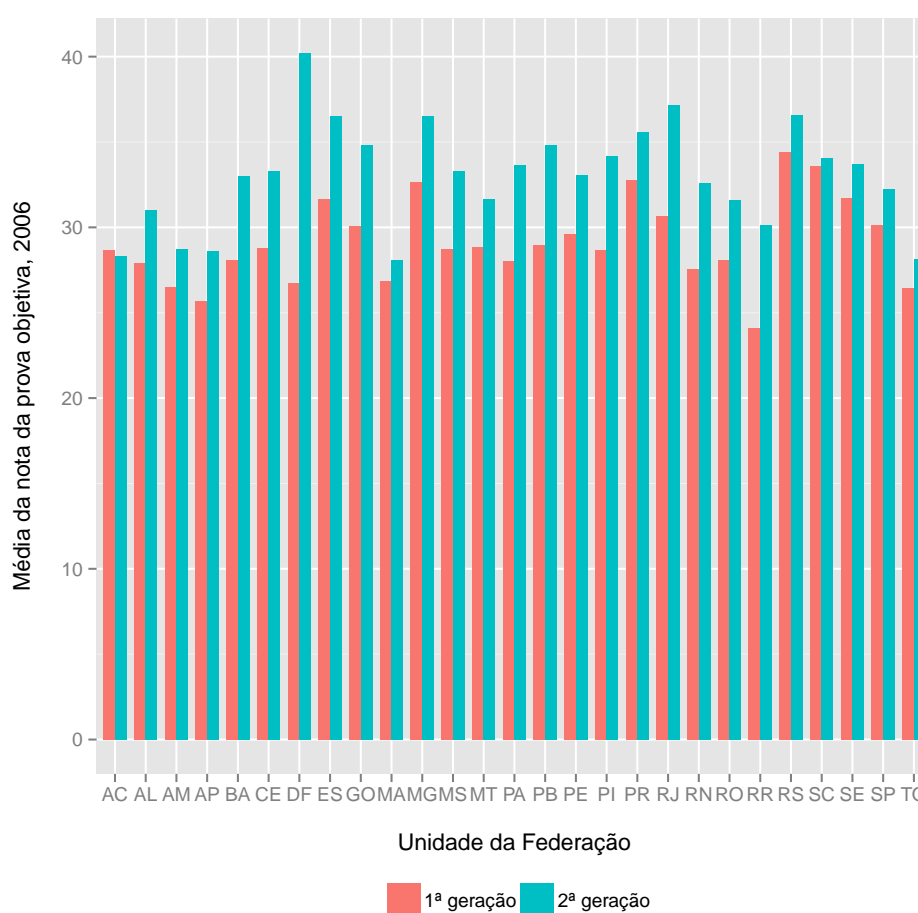
Tabela 2 – Estatística descritiva da nota da prova objetiva do ENEM 2006

	Nota da prova objetiva (2006)				
	Obs	Mínimo	Máximo	Média / Desvio-padrão	Mediana
1ª geração	11885	0	90,48	29,80 / 10,65	28,57
2ª geração	53027	0	93,65	33,60 / 13,07	30,16

Fonte: elaboração própria

A grande diferença de comportamento em relação aos salários aparece na comparação geográfica explicitada na figura 2. A diferença entre a maior média das notas da primeira e segunda gerações é de 5,8 pontos. Observa-se que em todas as UF, exceto o Acre, a segunda geração possui valor maior que a primeira, e no Distrito Federal essa diferença é a maior, alcançando 33,52%.

Figura 2 – Gráfico comparativo da média da nota da prova objetiva do ENEM por UF



Fonte: elaboração própria

Outra informação relevante é sobre a composição geográfica em 2006 e o comportamento migratório da amostra. A tabela 3 mostra que o estado de São Paulo apa-

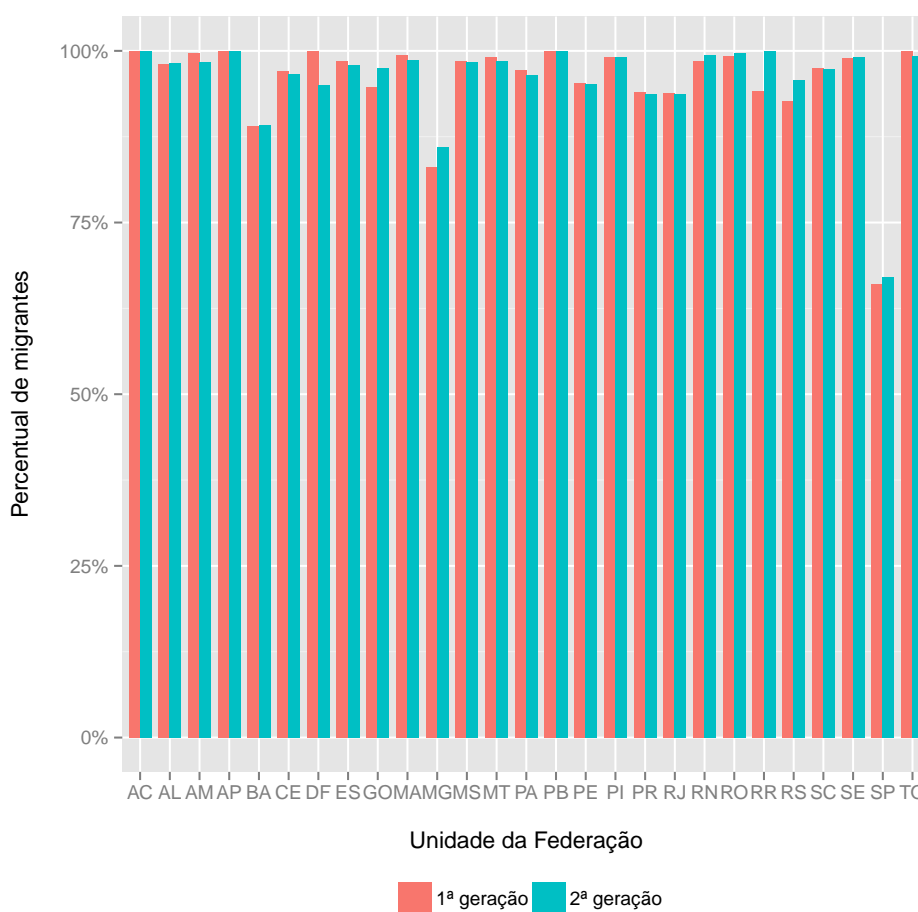
rece com a maior parcela da composição da amostra em ambas as gerações, porém a diferença entre os percentuais é grande, de 26,24 p.p.

Tabela 3 – Características da amostra sobre composição geográfica e migração

	Composição da amostra por UF		% de migração da amostra
	Maior percentual	Menor percentual	
1ª geração	SP (19,12%)	RR (0,14%)	88,86
2ª geração	SP (45,36%)	AC (0,18%)	81,77

Fonte: elaboração própria

Figura 3 – Gráfico comparativo do percentual de migrantes por UF



Fonte: elaboração própria

Grande parte da amostra migrou de UF entre 2006 e 2011, sendo que a primeira geração apresenta um percentual maior em relação à segunda. Ao analisar a migração por UF na figura 3, não há diferença geracional considerável (exceto em Roraima), mas sim diferença geográfica. Os estados com os menores índices são São

Paulo, Minas Gerais e Bahia. Os restantes possuíam taxas maior ou igual a 90%, com 6 UFs atingindo 100%.

Até agora foram apresentadas informações sobre quatro variáveis do modelo: $\log(\text{salário}_{2011})$, nota_{2006} , migrou e UF . Na tabela 9 do Apêndice encontram-se mais informações sobre as variáveis do modelo.

Das variáveis de características do indivíduo têm-se idade , que representa a idade do indivíduo, masc , uma *dummy* de sexo, branco , uma *dummy* de raça/cor, educ_mae_es e educ_pai_es , *dummies* que dizem se os pais do indivíduo possuem ensino superior, freq_esc , uma *dummy* que indica se o indivíduo está frequentando a escola em 2006, educ_11 , uma *dummy* que indica se o indivíduo possui onze anos de escolaridade e educ_12 , outra *dummy* que indica se o indivíduo possui doze ou mais anos de escolaridade.

Das variáveis de características de professores e diretores têm-se educ_prof_es , educ_prof_pos , educ_dir_es e educ_dir_pos , que mostram o total de professores ou diretores com ensino superior ou pós-graduação no município da escola em que o indivíduo estuda em 2006. As informações dos salários médios reais dos profissionais são postas em logaritmo e representadas pelas variáveis \ln_salmed_prof e \ln_salmed_dir .

Por fim, têm-se as variáveis de características da escola: esc_urb , *dummy* que indica se a escola se localiza em área urbana, esc_partic , *dummy* que indica se a escola pertence à rede particular e as *dummies* biblioteca , lab_cien e pc , que indicam se há presença de biblioteca, laboratório de ciências e computadores na escola.

Comparando os dados dessas variáveis entre as gerações, vê-se semelhanças entre os grupos, como a porcentagem de homens (43,6% da primeira geração e 46,1% da segunda) e de indivíduos que estudam em escola na área urbana (97,8% da primeira e 99% da segunda).

Porém as diferenças estão presentes: enquanto brancos compõe 32,1% da primeira geração, esse percentual sobe para 47,9% na segunda geração, uma diferença de 15,8 p.p. Indivíduos da primeira geração com pais com ensino superior mal chegam a 4% da amostra, enquanto cerca de 1 a cada 10 indivíduos da segunda geração possuem pais com ensino superior. A segunda geração também possui um percentual maior de indivíduos que estudavam em escola particular (19,5%), ao contrário dos quase 6% da primeira geração.

A infraestrutura das escolas também difere entre as gerações, que pode chegar a 14 p.p. A primeira geração possui mais indivíduos em escolas com biblioteca, porém a segunda geração conta com mais indivíduos em escolas mais modernas, com laboratórios de ciência e computadores. Quando se observa a frequência escolar, a

diferença chega a 22,1 p.p., com 91,1% da segunda geração frequentando a escola em 2006. Faz sentido a diferença ser tão grande, tendo em vista que esses indivíduos estão em idade de conclusão do Ensino Médio.

2.2 Metodologia aplicada

O trabalho se desenvolve a partir de um modelo primário, chamado de modelo original, e 4 modelos derivados do modelo original, em que se muda algumas variáveis. Para a estimação dos modelos, duas técnicas serão utilizadas: MQO e Variáveis Instrumentais, por meio do MQ2E. Todas as regressões dos modelos foram executadas com o *software* Stata®.

O foco é estimar empiricamente o efeito da qualidade da educação, representada pela nota da prova objetiva do ENEM 2006, sobre salário futuro, representado pelo salário médio real do indivíduo em 2011.

O modelo original a ser estimado é

$$\log(\text{salário}2011_{i,g}) = \varphi_0 + \varphi_1 \text{nota}2006_{i,g} + \varphi_2 I_{i,g} + \varphi_3 PD_{i,g} + \varphi_4 E_{i,g} + u_{i,g}$$

em que i e g são os índices que representam o indivíduo e sua geração, respectivamente $\text{salário}2011$ é o salário médio real em 2011, $\text{nota}2006$ é a nota da prova objetiva do ENEM 2006, I são as características do indivíduo, *background* familiar, além de controles de UF e migração, PD são as características dos professores e diretores e E são as características das escolas.

Os modelos derivados foram numerados de acordo com suas modificações. O modelo #1 utiliza a nota do ENEM 2006 e as variáveis monetárias em logaritmo. O modelo #2 utiliza a nota e as variáveis monetárias em nível. Os modelos #3 e #4 utilizam IV, assunto a ser tratado na subseção abaixo.

2.2.1 Variáveis Instrumentais

Variáveis instrumentais são utilizadas para eliminar a endogeneidade presente em uma variável explicativa, ou seja, instrumentalizar, e permitir que se observe uma variação exógena e seu efeito sobre a variável dependente do modelo. No modelo abordado no trabalho a variável $\text{nota}2006$ será instrumentalizada através da técnica de estimação conhecida como MQ2E, que consiste em 2 estágios de regressões por MQO:

1º estágio (equação reduzida)

$$\text{nota}2006_{i,g} = \pi_0 + \pi_1 I_{i,g} + \pi_2 PD_{i,g} + \pi_3 E_{i,g} + \pi_4 Z_{1i,g} + \pi_5 Z_{2i,g} + v_{1i,g}$$

para os modelos #1, #2 e #3 e

$$nota2006_{i,g} = \pi_0 + \pi_1 I_{i,g} + \pi_2 PD_{i,g} + \pi_3 E_{i,g} + \pi_4 Z_{i,g} + v_{1i,g}$$

para os modelos #4

2º estágio (equação estrutural)

$$\log(salário2011_{i,g}) = \varphi_0 + \varphi_1 \widehat{nota2006}_{i,g} + \varphi_2 I_{i,g} + \varphi_3 PD_{i,g} + \varphi_4 E_{i,g} + u_{i,g}$$

em que $\widehat{nota2006}$ são os valores preditos de $nota2006$ da equação reduzida, Z_1 representa med_prof_EI nos modelos #1 e #2 e $razmed_prof_pop_EI$ no modelo #3 e Z_2 sempre representa $ln_med_sal_EI$. Nos modelos #4, Z representa uma única IV, que é diferente para cada modificação do modelo.

Sobre as variáveis instrumentais, med_prof_EI representa o número médio de professores do ensino infantil do período em que o indivíduo estava no ensino infantil, $ln_med_sal_EI$ representa o logaritmo da média do salário médio real dos professores do ensino infantil do período em que o indivíduo estava no ensino infantil e $razmed_prof_pop_EI$ representa a média da razão do número de professores do ensino infantil e da população em idade de frequentar o ensino infantil no período em que o indivíduo estava no ensino infantil. Todas as variáveis são observadas a nível municipal e requerem que seja feita uma hipótese muito forte sobre os dados: o indivíduo não migrou de município entre o período em que fez ensino infantil e a prova do ENEM 2006.

A escolha das 3 variáveis instrumentais é embasada nos argumentos de que fazer o ensino infantil tem efeito positivo na probabilidade de concluir o ensino médio (CURI; MENEZES-FILHO, 2009) e é uma escolha feita no passado que tem efeito no presente, influenciada pela oferta de educação (SOUZA; CALDERINI, 2009).

3 Discussão e análise dos resultados

Os resultados das regressões estão ilustradas no Apêndice. Para os modelos regredidos por MQ2E, o resultado apresentado não inclui as regressões da equação estrutural. A variável *UF* indica se houve controle por *dummies* de UF na regressão. A ausência de R^2 em algumas regressões MQ2E se dá devido a uma fórmula de cálculo diferente do R^2 e subentende-se um grau de ajuste negativo, portanto considera-se zero.

O teste de restrição sobreidentificadora é realizado quando se tem mais IVs que variáveis explicativas endógenas. Com isso pode-se verificar se algum dos instrumentos está correlacionado com o erro. A hipótese nula do teste é que todas as IVs são exógenas. Antes de regredir os modelos é necessário realizar o teste. Ao realizá-lo nos modelos #1, #2 e #3, a hipótese nula não é rejeitada, ou seja, os instrumentos não são correlacionados com o erro estrutural e podem ser usados juntos.

Observa-se que nas tabelas 4, 5 e 6 a nota do ENEM só é significativa na regressão de MQO da primeira geração. No modelo original, o aumento da nota em 1 ponto causa, em média, um aumento aproximado do salário médio real em 0,168%. No modelo #1, o aumento da nota em 1% causa, em média, um aumento aproximado do salário médio real em 0,047%. No modelo #2, o aumento da nota em 1 ponto causa, em média, um aumento aproximado de R\$ 4,48 no salário médio real esperado.

O uso de *razmed_prof_pop_EI* substituindo *med_prof_EI* como IV no modelo #3 não implica mudanças significativas, como mostra a tabela 7. Nos modelos #1 e #2 *ln_salmed_prof* é a única variável significativa do modelo, exceto *migrou*, a 10%. Já no modelo #3 as variáveis *idade*, *educ_prof_es*, *freq_esc* e *educ_12* se tornam significantes a 10% na regressão da primeira geração. Na regressão da segunda geração, impacta apenas *ln_salmed_prof*, que também passa a ser significativa a 10%. O modelo #4, na tabela 8, não obteve êxito em mostrar diferença nos resultados comparados aos modelos anteriores.

A variável indicadora de migração, porém, foi a única a ser fortemente significativa em todos os modelos (exceto nas regressões da primeira geração do MQ2E do modelo #2 e do MQ2E com *ln_med_sal_EI* do modelo #4), com p-valor abaixo de 0,01. Mas seu sinal negativo contradiz o esperado pela literatura, uma vez que pessoas que migram estão a procura de melhores empregos, com remunerações maiores (RIOS-NETO, 2005).

Conclusão

A meta do trabalho foi trazer o contexto da análise da relação entre qualidade da educação e salários futuros de Curi e Menezes-Filho (2005) utilizando um *cross-section* de dados defasados identificados para o mesmo indivíduo e a metodologia econométrica de variáveis instrumentais.

A falta de resultados significativos e que corroborem com a literatura indica que o problema pode estar na base de dados, uma vez que o período de 5 anos entre o ENEM e a inserção do indivíduo no mercado de trabalho é muito pequeno e não estaria captando muito bem as pessoas que entram no mercado de trabalho após concluírem o ensino superior. Outros métodos de estimação também podem prover resultados melhores, como a regressão quantílica, que está fora do escopo do presente trabalho.

Referências

- BARRO, R. J. Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, v. 106, n. 2, p. 407–441, 1991.
- BARRO, R. J.; LEE, J.-W. Schooling quality in a cross section of countries. *Economica*, v. 68, n. 272, p. 465–488, 2001.
- BISHOP, J. Is the test score decline responsible for the productivity growth decline? *The American Economic Review*, v. 79, n. 1, p. 178–197, 1989.
- CHETTY, R.; FRIEDMAN, J. N.; HILGER, N.; SAEZ, E.; SCHAZENBACH, D. W.; YAGAN, D. *How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR*. Cambridge, Estados Unidos, 2011. Disponível em: <<http://www.nber.org/papers/w16381>>.
- CHETTY, R.; FRIEDMAN, J. N.; ROCKOFF, J. E. Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *The American Economic Review*, v. 104, n. 9, p. 2593–2632, 2014.
- _____. Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *The American Economic Review*, v. 104, n. 9, p. 2633–2779, 2014.
- CURI, A. Z.; MENEZES-FILHO, N. A. A relação entre o desempenho escolar e salários no Brasil. In: ENCONTRO NACIONAL DE ECONOMIA, 33., 2005, Natal. *Anais Eletrônicos...* 2005. Disponível em: <[http://www.anpec.org.br/encontro2005-artigos/A05A158.pdf](http://www.anpec.org.br/encontro2005/artigos/A05A158.pdf)>.
- _____. A relação entre educação pré-primária, salários, escolaridade e proficiência escolar no Brasil. *Estudos Econômicos*, v. 39, n. 4, p. 811–850, 2009.
- FELÍCIO, F.; FERNANDES, R. O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental no estado de São Paulo. In: ENCONTRO NACIONAL DE ECONOMIA, 33., 2005, Natal. *Anais Eletrônicos...* 2005. Disponível em: <<http://www.anpec.org.br/encontro2005/artigos/A05A157.pdf>>.
- HANUSHEK, E. A. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, v. 24, n. 3, p. 1141–1177, 1986.
- _____. The failure of input-based schooling policies. *The Economic Journal*, v. 113, n. 485, p. F64–F98, 2003.
- _____. Economic growth in developing countries: The role of human capital. *Economics of Education Review*, v. 37, p. 204–212, 2013.
- HANUSHEK, E. A.; KIMKO, D. D. Schooling, labor force quality, and the growth of nations. *The American Economic Review*, v. 90, n. 5, p. 1184–1208, 2000.
- JONES, C. O Modelo de Solow. In: _____. *Introdução à Teoria do Crescimento Econômico*. Rio de Janeiro: Campus, 2000. cap. 2, p. 16–43.

MINCER, J. Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, v. 66, n. 4, p. 281–302, 1958.

_____. Human Capital and Economic Growth. *Economics of Education Review*, v. 3, n. 3, p. 195–205, 1984.

MURNANE, R.; WILLET, J.; LEVY, F. The growing importance of cognitive skill in wage determination. *Review of Economics and Statistics*, v. 77, n. 2, p. 251–266, 1995.

PSACHAROPOULOS, G.; WOODHALL, M. Internal Efficiency and Educational Quality. In: _____. *Education for Development: An Analysis of Investment Choices*. Nova Iorque: Oxford University Press, 1985. cap. 8, p. 205–243.

RIOS-NETO, E. L. G. Questões emergentes na análise demográfica: o caso brasileiro. *Revista Brasileira de Estudos de População*, v. 22, n. 2, p. 371–408, 2005.

ROMER, P. M. Endogenous Technological Change. *Journal of Political Economy*, v. 98, n. 5, p. S71–S102, 1990.

SCHULTZ, T. W. Capital Formation by Education. *Journal of Political Economy*, v. 68, n. 6, p. 571–583, 1960.

SOLOW, R. M. A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics* 1, v. 70, n. 1, p. 65–94, 1956.

SOUZA, A. P.; CALDERINI, S. R. Pré-escola no Brasil: Seu Impacto na Qualidade da Educação Fundamental. In: ENCONTRO NACIONAL DE ECONOMIA, 37., 2009, Foz do Iguaçu. *Anais Eletrônicos...* 2009. Disponível em: <[http://www.anpec.org-br/encontro2009/inscricao.on/arquivos/000-2faec48de6ce147ca713e5bc1a22e59f-.docx](http://www.anpec.org.br/encontro2009/inscricao.on/arquivos/000-2faec48de6ce147ca713e5bc1a22e59f-.docx)>.

Apêndices

APÊNDICE A – Regressões

Tabela 4 – Estimação dos modelos originais

Variáveis	Variável dependente: Logaritmo do salário real médio			
	MQO		MQ2E	
	1ª Geração	2ª Geração	1ª Geração	2ª Geração
nota_prova	0,00168** (0,00071)	-0,00031 (0,00026)	0,03413 (0,05921)	0,00476 (0,01052)
idade	-0,01261 (0,00852)	-0,00114 (0,00424)	0,00292 (0,02753)	0,00751 (0,02150)
masc	-0,00354 (0,01352)	-0,00274 (0,00587)	-0,07919 (0,14794)	-0,01757 (0,03421)
branco	-0,01544 (0,01505)	-0,00509 (0,00619)	-0,05361 (0,08283)	-0,01839 (0,02267)
educ_mae_es	0,00748 (0,03744)	-0,01293 (0,01019)	-0,09687 (0,18928)	-0,03267 (0,04222)
educ_pai_es	0,03857 (0,03857)	0,01710* (0,01027)	-0,07869 (0,21636)	-0,00137 (0,03990)
educ_prof_es	-0,00005* (0,00003)	-0,00000 (0,00001)	-0,00004 (0,00005)	-0,00002 (0,00003)
educ_prof_pos	0,00023 (0,00169)	0,00015 (0,00087)	0,00141 (0,00257)	-0,00036 (0,00105)
educ_dir_es	-0,00027 (0,00181)	0,00020 (0,00072)	-0,00079 (0,00330)	0,00019 (0,00077)
educ_dir_pos	0,04294 (0,04497)	-0,01103 (0,01767)	0,06027 (0,08723)	-0,00147 (0,02231)
ln_salmed_prof	0,02604 (0,02569)	0,01195 (0,01169)	-0,03309 (0,07980)	0,02263* (0,01241)
ln_salmed_dir	-0,01033 (0,01462)	0,00078 (0,00598)	-0,02119 (0,03451)	-0,00257 (0,00719)
esc_urb	-0,06054 (0,05499)	0,00425 (0,04346)	-0,06980 (0,06878)	-0,00973 (0,04697)
esc_partic	-0,07077** (0,03038)	0,01118 (0,00862)	-0,35491 (0,51532)	-0,05281 (0,13015)
biblioteca	0,01467 (0,01845)	0,00645 (0,00768)	-0,00532 (0,05007)	0,00156 (0,01292)
lab_cien	0,00395 (0,01670)	-0,00491 (0,00879)	-0,03451 (0,07836)	-0,00996 (0,01526)
pc	0,01098 (0,01475)	-0,00520 (0,00663)	0,01775 (0,02218)	-0,01030 (0,01224)
freq_esc	0,13823** (0,06935)	-0,00727 (0,03495)	0,10833 (0,09575)	-0,01509 (0,03851)
educ_11	0,09807 (0,07341)	0,00705 (0,04201)	-0,08294 (0,32762)	-0,02656 (0,08499)
educ_12	0,13105* (0,07098)	0,00473 (0,03722)	-0,12680 (0,47709)	-0,04752 (0,11325)
migrou	-0,07650*** (0,02239)	-0,15612*** (0,00793)	-0,08035*** (0,02711)	-0,15817*** (0,00802)
UF	SIM	SIM	SIM	SIM
Constante	6,98713*** (0,26606)	6,99092*** (0,12106)	6,42901*** (1,20405)	6,68808*** (0,59697)
Observações	6,587	33,767	6,287	32,897
R ²	0,01104	0,01298		0,00250

Erro-padrão entre parênteses

*** p<0.01, ** p<0.05, * p<0.1

Tabela 5 – Estimação dos modelos derivados #1

Variáveis	Variável dependente: Logaritmo do salário real médio			
	MQO		MQ2E	
	1ª Geração	2ª Geração	1ª Geração	2ª Geração
ln_nota_prova	0,04658** (0,02186)	-0,01493 (0,00924)	1,07524 (1,68330)	0,19534 (0,39094)
idade	-0,01279 (0,00854)	-0,00153 (0,00424)	0,00028 (0,02162)	0,00959 (0,02386)
masc	-0,00240 (0,01351)	-0,00251 (0,00586)	-0,07162 (0,12219)	-0,01785 (0,03167)
branco	-0,01517 (0,01507)	-0,00489 (0,00619)	-0,05379 (0,07541)	-0,01952 (0,02283)
educ_mae_es	0,00879 (0,03742)	-0,01250 (0,01018)	-0,07203 (0,13437)	-0,03209 (0,03780)
educ_pai_es	0,04097 (0,03855)	0,01735* (0,01026)	-0,04946 (0,15254)	0,00048 (0,03318)
educ_prof_es	-0,00005* (0,00003)	-0,00000 (0,00001)	-0,00003 (0,00006)	-0,00002 (0,00003)
educ_prof_pos	0,00019 (0,00169)	0,00019 (0,00087)	0,00078 (0,00204)	-0,00029 (0,00100)
educ_dir_es	-0,00008 (0,00182)	0,00022 (0,00072)	-0,00020 (0,00273)	0,00022 (0,00077)
educ_dir_pos	0,03665 (0,04520)	-0,01223 (0,01768)	0,03349 (0,06171)	-0,00299 (0,02119)
ln_salmed_prof	0,02559 (0,02571)	0,01195 (0,01169)	-0,03100 (0,06925)	0,02268* (0,01243)
ln_salmed_dir	-0,00889 (0,01464)	0,00058 (0,00598)	-0,01859 (0,03057)	-0,00308 (0,00730)
esc_urb	-0,06221 (0,05501)	0,00396 (0,04346)	-0,08277 (0,06707)	-0,01017 (0,04686)
esc_partic	-0,06609** (0,03011)	0,01162 (0,00851)	-0,28711 (0,35926)	-0,05526 (0,12189)
biblioteca	0,01561 (0,01846)	0,00665 (0,00768)	0,00076 (0,03808)	0,00078 (0,01355)
lab_cien	0,00462 (0,01673)	-0,00453 (0,00879)	-0,03496 (0,07299)	-0,01111 (0,01656)
pc	0,01110 (0,01477)	-0,00488 (0,00664)	0,01648 (0,02031)	-0,01106 (0,01305)
freq_esc	0,13760** (0,06937)	-0,00697 (0,03495)	0,10040 (0,09997)	-0,01831 (0,04079)
educ_11	0,09880 (0,07344)	0,00817 (0,04202)	-0,09227 (0,31204)	-0,03627 (0,09601)
educ_12	0,13284* (0,07099)	0,00545 (0,03722)	-0,12267 (0,42552)	-0,05624 (0,11916)
migrou	-0,07724*** (0,02240)	-0,15629*** (0,00794)	-0,08560*** (0,02627)	-0,15846*** (0,00804)
UF	SIM	SIM	SIM	SIM
Constante	6,87737*** (0,27392)	7,03940*** (0,12608)	3,90247 (5,00869)	6,14955*** (1,61892)
Observações	6,574	33,745	6,275	32,875
R ²	0,01080	0,01304		

Erro-padrão entre parênteses

*** p<0.01, ** p<0.05, * p<0.1

Tabela 6 – Estimação dos modelos derivados #2

Variáveis	Variável dependente: Salário real médio, R\$			
	MQO		MQ2E	
	1ª Geração	2ª Geração	1ª Geração	2ª Geração
nota_prova	4,47685*** (1,36419)	-0,71808 (0,48804)	125,24091 (115,49018)	-12,87762 (20,39475)
idade	-3,05352 (16,47229)	-0,38081 (7,98067)	57,04223 (62,37537)	-24,49108 (41,22072)
masc	9,27443 (26,13176)	-1,78667 (11,06380)	-301,51292 (299,65257)	37,00550 (65,99560)
branco	-25,59325 (29,05651)	-10,45674 (11,66556)	-176,53810 (150,54353)	14,24482 (43,05704)
educ_mae_es	-20,59192 (71,58988)	-26,88595 (19,19602)	-396,86360 (374,92075)	20,58099 (81,91144)
educ_pai_es	43,50317 (74,54640)	14,29929 (19,35067)	-385,26980 (424,43638)	59,58623 (78,40304)
educ_prof_es	-0,11680** (0,05823)	0,01129 (0,02615)	-0,09864 (0,08751)	0,04410 (0,06101)
educ_prof_pos	0,93234 (3,26479)	-0,24420 (1,67127)	4,90151 (6,12627)	0,35972 (1,96612)
educ_dir_es	-0,26914 (3,57001)	-0,19443 (1,41750)	-3,98678 (6,34743)	-0,45232 (1,49336)
educ_dir_pos	86,76373 (87,97034)	-6,17006 (33,85216)	224,04756 (184,45263)	-21,04727 (42,27957)
salmed_prof	0,02653 (0,03651)	0,00983 (0,01430)	-0,08859 (0,12251)	0,01222 (0,01496)
salmed_dir	-0,00187 (0,00751)	0,00105 (0,00252)	-0,01145 (0,01436)	0,00191 (0,00291)
esc_urb	-90,03500 (104,35451)	-46,77852 (80,90089)	-74,79893 (154,40422)	-49,14794 (81,67609)
esc_partic	-79,49554 (58,76492)	24,91565 (16,23873)	-1,142,42299 (1,020,03402)	173,80896 (250,19708)
biblioteca	52,53148 (35,67078)	6,23039 (14,47670)	-45,27288 (107,26891)	18,79232 (25,62766)
lab_cien	-5,36139 (32,32849)	-20,42034 (16,55999)	-154,80806 (150,62431)	-6,33772 (28,92136)
pc	20,90873 (28,47798)	-5,82836 (12,49107)	38,71183 (45,27087)	6,41691 (24,08819)
freq_esc	234,89320* (134,73879)	44,41143 (65,55046)	130,63767 (222,10052)	63,00107 (73,08112)
educ_11	120,00149 (142,59745)	57,99950 (78,82628)	-560,02487 (683,31527)	140,61436 (159,71179)
educ_12	221,15443 (137,87535)	80,59834 (69,83734)	-740,56280 (941,73867)	202,12386 (215,59692)
migrou	-131,66245*** (43,26907)	-191,59811*** (14,94776)	-114,77453* (65,74962)	-191,08540*** (15,09762)
UF	SIM	SIM	SIM	SIM
Constante	995,47400** (392,04977)	1,381,98716*** (179,71811)	-2,872,69995 (3,743,49916)	2,079,87760* (1,184,15525)
Observações	6,660	34,063	6,660	34,063
R ²	0,01140	0,00645		

Erro-padrão entre parênteses

*** p<0.01, ** p<0.05, * p<0.1

Tabela 7 – Estimação dos modelos derivados #3

Variável dependente: Logaritmo do salário real médio		
Variáveis	MQ2E	
	1ª Geração	2ª Geração
nota_prova	-0,04258 (0,03309)	-0,00365 (0,00654)
idade	-0,03031* (0,01804)	-0,00933 (0,01376)
masc	0,11138 (0,08400)	0,00937 (0,02174)
branco	0,05124 (0,04917)	-0,00097 (0,01490)
educ_mae_es	0,14173 (0,11353)	0,00008 (0,02742)
educ_pai_es	0,19564 (0,12797)	0,02944 (0,02606)
educ_prof_es	-0,00008* (0,00004)	0,00001 (0,00002)
educ_prof_pos	-0,00075 (0,00234)	0,00010 (0,00095)
educ_dir_es	0,00245 (0,00274)	0,00002 (0,00075)
educ_dir_pos	-0,02904 (0,07013)	-0,01191 (0,01978)
ln_salmed_prof	0,06150 (0,05397)	0,02345* (0,01235)
ln_salmed_dir	0,01744 (0,02531)	0,00044 (0,00654)
esc_urb	-0,09260 (0,07356)	-0,01712 (0,04627)
esc_partic	0,31114 (0,28993)	0,05105 (0,08109)
biblioteca	0,05307 (0,03473)	0,00982 (0,01005)
lab_cien	0,06380 (0,04751)	-0,00009 (0,01179)
pc	-0,00025 (0,02049)	-0,00213 (0,00924)
freq_esc	0,17757* (0,09208)	-0,00273 (0,03646)
educ_11	0,32719 (0,19965)	0,03227 (0,06236)
educ_12	0,48225* (0,27744)	0,03791 (0,07620)
migrou	-0,09003*** (0,02887)	-0,15808*** (0,00799)
UF	SIM	SIM
Constante	7,93256*** (0,73780)	7,15476*** (0,38346)
Observações	6,287	32,897
R^2		0,00848

Erro-padrão entre parênteses

*** p<0.01, ** p<0.05, * p<0.1

Tabela 8 – Estimação dos modelos derivados #4

Variáveis	Variável dependente: Logaritmo do salário real médio					
	IV: <i>med_prof_EI</i>		IV: <i>ln_med_sal_EI</i>		IV: <i>razmed_prof_pop_EI</i>	
	1ª Geração	2ª Geração	1ª Geração	2ª Geração	1ª Geração	2ª Geração
nota_prova	0,04795 (0,06764)	0,02886 (0,16371)	-0,21406 (0,63166)	0,00383 (0,01077)	-0,01066 (0,02751)	-0,00529 (0,00655)
idade	0,00977 (0,03449)	0,05709 (0,32685)	-0,10459 (0,27572)	0,00565 (0,02199)	-0,01835 (0,01575)	-0,01119 (0,01374)
masc	-0,12307 (0,17555)	-0,09542 (0,52019)	0,53742 (1,57018)	-0,01459 (0,03501)	0,02789 (0,07229)	0,01294 (0,02157)
branco	-0,07512 (0,08933)	-0,06430 (0,33240)	0,28567 (0,86553)	-0,01647 (0,02317)	0,00023 (0,03833)	0,00483 (0,01470)
educ_mae_es	-0,12968 (0,20613)	-0,12754 (0,64333)	0,67514 (1,97034)	-0,02905 (0,04317)	0,04470 (0,09045)	0,00674 (0,02759)
educ_pai_es	-0,12823 (0,24873)	-0,09075 (0,60537)	0,80895 (2,26407)	0,00203 (0,04078)	0,08359 (0,10700)	0,03532 (0,02635)
educ_prof_es	-0,00003 (0,00005)	-0,00008 (0,00043)	-0,00019 (0,00040)	-0,00001 (0,00003)	-0,00006* (0,00003)	0,00001 (0,00002)
educ_prof_pos	0,00143 (0,00278)	-0,00153 (0,00946)	-0,00558 (0,01899)	-0,00031 (0,00106)	-0,00005 (0,00185)	0,00041 (0,00095)
educ_dir_es	-0,00192 (0,00335)	0,00090 (0,00406)	0,00969 (0,02766)	0,00017 (0,00077)	0,00015 (0,00209)	0,00008 (0,00074)
educ_dir_pos	0,09386 (0,09415)	0,02273 (0,19064)	-0,22871 (0,75747)	-0,00262 (0,02247)	0,02980 (0,05473)	-0,01687 (0,01933)
ln_salmed_prof	-0,03268 (0,09192)	0,00418 (0,04569)	0,27296 (0,78654)	0,02272* (0,01239)	0,04188 (0,04412)	0,01380 (0,01187)
ln_salmed_dir	-0,03454 (0,04004)	-0,00829 (0,05137)	0,10380 (0,32351)	-0,00223 (0,00723)	-0,00385 (0,02051)	0,00198 (0,00636)
esc_urb	-0,05808 (0,07062)	0,01313 (0,07129)	-0,14356 (0,29342)	-0,01055 (0,04693)	-0,06082 (0,05608)	0,00283 (0,04371)
esc_partic	-0,47534 (0,59254)	-0,34596 (2,00439)	1,80015 (5,48597)	-0,04133 (0,13326)	0,03492 (0,24067)	0,07228 (0,08061)
biblioteca	-0,02349 (0,06059)	-0,02310 (0,16604)	0,18360 (0,48646)	0,00248 (0,01311)	0,02506 (0,02951)	0,01116 (0,01017)
lab_cien	-0,05422 (0,08768)	-0,03798 (0,18590)	0,28359 (0,81226)	-0,00887 (0,01549)	0,01942 (0,03820)	0,00024 (0,01148)
pc	0,01922 (0,02242)	-0,03500 (0,16741)	-0,04048 (0,15939)	-0,00939 (0,01243)	0,00795 (0,01583)	0,00023 (0,00948)
freq_esc	0,09737 (0,10714)	-0,05292 (0,25949)	0,33237 (0,63064)	-0,01372 (0,03860)	0,14942** (0,07485)	0,00050 (0,03657)
educ_11	-0,16387 (0,39423)	-0,19285 (1,12297)	1,24408 (3,38939)	-0,02006 (0,08648)	0,16624 (0,17158)	0,04242 (0,06184)
educ_12	-0,23834 (0,54747)	-0,28962 (1,65249)	1,84384 (5,02289)	-0,03807 (0,11564)	0,22978 (0,23141)	0,05534 (0,07586)
migrou	-0,07336** (0,02908)	-0,15710*** (0,01080)	-0,11168 (0,11905)	-0,15816*** (0,00800)	-0,07732*** (0,02291)	-0,15598*** (0,00797)
UF	SIM	SIM	SIM	SIM	SIM	SIM
Constante	6,06545*** (1,38957)	5,41811 (8,82799)	11,29391 (12,42895)	6,73966*** (0,61059)	7,22701*** (0,60680)	7,26076*** (0,37356)
Observações	6,587	33,767	6,287	32,897	6,581	33,734
R^2				0,00621		0,00226

Erro-padrão entre parênteses

*** p<0.01, ** p<0.05, * p<0.1

APÊNDICE B – Estatísticas descritivas das variáveis

Tabela 9 – Estatísticas descritivas das variáveis, por geração

Variáveis	Observações		Média		Desvio-Padrão	
	1ª Geração	2ª Geração	1ª Geração	2ª Geração	1ª Geração	2ª Geração
ln_salmed11	11763	52607	6,941	6,937	0,534	0,537
nota_prova	11885	53027	29,804	33,600	10,648	13,068
idade	11885	53027	20,732	17,765	0,786	0,743
masc	11885	53027	0,436	0,461	0,496	0,498
branco	11885	53027	0,321	0,479	0,467	0,500
educ_mae_es	11885	53027	0,037	0,101	0,188	0,302
educ_pai_es	11885	53027	0,027	0,095	0,162	0,293
educ_prof_es	9407	44927	374,724	457,313	639,769	691,136
educ_prof_pos	9407	44927	4,695	6,597	8,873	10,526
educ_dir_es	6922	34800	19,191	26,183	31,187	37,071
educ_dir_pos	6922	34800	0,568	0,794	1,021	1,223
ln_salmed_prof	9397	44900	7,128	7,280	0,614	0,596
ln_salmed_dir	6910	34779	8,283	8,412	0,708	0,720
esc_urb	11885	53027	0,978	0,990	0,147	0,099
esc_partic	11885	53027	0,052	0,195	0,223	0,396
biblioteca	11885	53027	0,731	0,645	0,443	0,478
lab_cien	11885	53027	0,676	0,819	0,468	0,385
pc	11885	53027	0,497	0,622	0,500	0,485
freq_esc	11885	53027	0,684	0,911	0,465	0,285
educ_11	11885	53027	0,083	0,019	0,276	0,138
educ_12	11885	53027	0,223	0,063	0,416	0,243
migrou	11885	53027	0,889	0,818	0,315	0,386
med_prof_EI	11885	53027	368,215	566,131	867,585	1082,152
ln_med_sal_EI	10131	47949	6,475	6,684	0,714	0,697
razmed_prof_pop_EI	11814	52822	0,002	0,002	0,002	0,002